

# EXHIBIT E

**SUSMAN GODFREY L.L.P.**

A REGISTERED LIMITED LIABILITY PARTNERSHIP  
ONE MANHATTAN WEST  
NEW YORK, NEW YORK 10001-8602  
(212) 336-8330  
FAX (212) 336-8340  
WWW.SUSMANGODFREY.COM

---

SUITE 5100  
1000 LOUISIANA STREET  
HOUSTON, TEXAS 77002-5096  
(713) 651-9366

---

SUITE 1400  
1900 AVENUE OF THE STARS  
LOS ANGELES, CALIFORNIA 90067-6029  
(310) 789-3100

---

SUITE 3000  
401 UNION STREET  
SEATTLE, WASHINGTON 98101-2683  
(206) 516-3880

---

ZACH SAVAGE  
DIRECT DIAL (212) 729-2022

E-MAIL ZSAVAGE@SUSMANGODFREY.COM

July 31, 2024

**VIA ELECTRONIC MAIL**

Vera Ranieri  
Morrison & Foerster LLP  
vranieri@mofo.com  
425 Market Street  
San Francisco, CA 94105-2482

Elana Nightingale Dawson  
Latham & Watkins LLP  
elana.nightingaledawson@lw.com  
505 Montgomery Street, Suite 2000  
San Francisco, CA 94111

Katie Lynn Joyce  
Keker, Van Nest & Peters LLP  
kjoyce@keker.com  
633 Battery Street  
San Francisco, California 94111-1809

Re: *The New York Times Company v. Microsoft Corporation, et al.*,  
Case No.: 23-cv-11195-SHS

Counsel:

I write regarding OpenAI's Responses and Objections to The Times's Second Set of Requests for Production of Documents, served on July 12, 2024 (the "R&Os"). Please offer a time to meet and confer about these responses no later than August 6.

## I. Global Issues

### a. Offering to Meet and Confer in Lieu of Substantively Responding

For 26 of the 57 Requests in this set of RFPs, *almost half*, OpenAI merely offers to meet and confer instead of stating whether it will or will not produce documents. This wide-scale refusal to take a position violates both the letter and spirit of the Federal Rules. While meeting-and-confering is an important part of this process, it should not be used as a substitute for substantively responding to Requests, like these, that are clear on their face. These RFPs seek discovery about plainly relevant topics, including:

- The Training Datasets for Defendants’ Generative AI Models, including Defendants’ selection and prioritization of content for these Datasets, Defendants’ decision not to publicly release these Datasets, and the removal of any content from these Datasets (RFPs 16, 18, 19, 25, 29, 33, 56, 59).
- The functioning of Defendants’ Generative AI Products and Services, including changes to Defendants’ Products and Services as well as technical documentation describing the Products and Services (RFPs 35, 37, 38, 46).
- Problematic behaviors exhibited by Defendants’ Generative AI Models, Products, and Services, including Defendants’ knowledge of and efforts to address memorization, hallucinations, and verbatim copies of copyrighted content, as well as communications with other LLM developers about these issues (RFPs 39, 40, 41, 57, 63).
- User engagement with Defendants’ Generative AI Products and Services, including logs that track and/or analyze user queries, user subscription information, click-through data, and data regarding the extent to which Defendants’ Products and Services direct users to Times content (RFPs 45, 47, 48, 51, 52).
- Defendants’ revenues and business plans, including revenues attributable to Defendants’ generative search products as well as Defendants’ plans for future products (RFPs 64, 67, 68, 72).

Prior to the meet and confer, for the above-listed RFPs, The Times requests that OpenAI reconsider its non-committal responses and commit to producing documents. Doing so will enable the parties to spend the time efficiently to address any questions that should be discussed.

### b. Limitation to ChatGPT and “Only the Models Used for ChatGPT”

OpenAI improperly defines “Generative AI Products and Services” as limited to ChatGPT. R&Os at 5. To the extent that OpenAI seeks to exclude Browse with Bing or the APIs, plugins, and Custom GPTs that provide access to Generative AI Models from this definition, such exclusion is improper.

In addition, The Times objects to OpenAI’s apparent intention to exclude products and services that OpenAI considers to be Microsoft products and services, such as Copilot. *See, e.g.*, R&Os at 5, 36, 38, 41, 42. When The Times previously moved to compel on this issue, OpenAI promptly withdrew this discovery limitation, agreeing to “conduct a reasonable search for non-

custodial documents concerning Bing Chat and Copilot.” Dkt. 136-1 at 1-2. OpenAI should agree to do the same for these RFPs.

OpenAI also improperly defines “AI Models”, “Generative AI Models”, and “Text Generation AI Models” as limited to the “models used for ChatGPT.” R&Os at 5. As you know from the parties’ motion practice on this issue, The Times’s position is that all of Defendants’ models are relevant and the proper subject of discovery. Dkt. 128 at 1-2. Please confirm that, if The Times prevails on that motion to compel, OpenAI will withdraw this limitation for these RFPs.

### **c. Improper Use of “Sufficient to Show”**

For many RFPs, OpenAI has improperly limited its production to documents “sufficient to show” the requested information (RFPs 24, 26, 27, 28, 32, 42, 43, 44, 48, 58, 70). This self-imposed “sufficient-to-show” limitation is improper because that approach “would let [OpenAI] decide what the answer[s] [are] on the merits [of factual issues in the case] and then limit discovery accordingly.” *In re Stubhub Refund Litig.*, 2022 WL 1640304, at \*2 (N.D. Cal. May 24, 2022). OpenAI should also search for and produce custodial documents in response to these requests.

## **II. Issues with Specific RFPs**

### **a. Requests for Which OpenAI Refuses to Produce Documents**

For seven requests, OpenAI refuses to produce any documents. Each of these requests seeks discovery into relevant information, and OpenAI should search for and produce the requested documents.

RFP 53 seeks documents “concerning the decision to suspend Browse with Bing and whether or not to suspend Bing Chat.” Contrary to OpenAI’s assertion, R&Os at 36, this request is not “vague and ambiguous.” But for the avoidance of doubt, The Times seeks information about OpenAI’s publicly reported decision in July 2023 to suspend Browse with Bing due to concerns with “the feature inadvertently allow[ing] users to bypass paywalls.”<sup>1</sup> Such documents are relevant to, among other issues, The Times’s contributory infringement claim as well as The Times’s allegations about how Defendants’ products are harming The Times. Compl. ¶¶ 154-57.

RFP 54 seeks documents “concerning any public statements made by You or Your employees relating to this litigation, including: . . . Your claim that the New York Times ‘hacked’ Defendants’ Generative AI Products and Services.” OpenAI said it will produce documents “on which it will rely at summary judgment or at trial at the appropriate time.” R&Os at 37. That response adds nothing because OpenAI should already be producing any documents it will rely on at summary judgment or trial. The Times is entitled to explore the basis for OpenAI’s stated defenses in this case, including to discover whether OpenAI’s internal documents or communications undermine its public assertions. *See, e.g.*, Dkt. 52 (OpenAI’s MTD) at 2 (arguing that The Times “paid someone to hack OpenAI’s products” and claiming this “truth [] will come out in the course of this case”).

---

<sup>1</sup> *OpenAI Pauses Bing Search Feature over Paywall Bypass Abilities*, THE REGISTER (July 5, 2023), [https://www.theregister.com/2023/07/05/openai\\_pauses\\_bing\\_search](https://www.theregister.com/2023/07/05/openai_pauses_bing_search).

RFP 55 seeks documents “concerning any firings, terminations, lay-offs, impositions of mandatory leave, disciplinary actions, departures, resignations, or other personnel-related actions Defendants have taken or considered concerning individuals or teams involved with any ethics, harm reduction safety, lawfulness, and responsible AI initiatives.” OpenAI’s objection that the Request is “vague and ambiguous” is unfounded, including because The Times identified, by way of example, specific employees and board members to whom this request applies. R&Os at 37-38. The requested documents are relevant to The Times’s allegation that Defendants’ infringement is willful, Compl. ¶¶ 124-26, particularly considering the numerous, highly publicized reports of terminations and resignations over issues including AI safety and ethics,<sup>2</sup> which are squarely implicated in this case.

RFP 61 seeks documents “concerning any indemnification obligations, common interest agreements, or joint defense agreements between Defendants related to Defendants’ Generative AI Products or Services.” The requested documents are relevant to, among other issues, The Times’s vicarious and contributory infringement claims. Although OpenAI has asserted that some of these documents are privileged, that does not relieve you of your obligation to search for and produce all responsive, non-privileged documents and log any withheld documents. Whether this request is “duplicative” to discovery served on Microsoft is of no moment. R&Os at 42. OpenAI cannot withhold relevant documents on the ground that Microsoft *may* produce those same documents. In any event, OpenAI will likely have unique documents responsive to this request, including internal communications among OpenAI employees.

RFP 65 seeks documents concerning “Your public relations strategies as they pertain to Generative AI Models, Generative AI Products and Services, or copyright.” OpenAI’s assertion that this Request seeks “publicly available” information is misguided. R&Os at 45. The Times seeks nonpublic documents and communications about OpenAI’s public relations strategies, which are relevant to, among other issues, willfulness, fair use and the substitutive effect of Defendants’ products on journalism. For example, if OpenAI has considered marketing ChatGPT as a substitute for news media, documents discussing those plans would be squarely relevant.

RFP 66 seeks documents “concerning fundraising efforts related to Defendants’ Generative AI Products and Services, including any investor solicitation efforts and any solicitation of charitable or in-kind contributions.” Contrary to OpenAI’s assertion, R&Os at 46, these documents are relevant, including for purposes of quantifying Defendants’ unjust profits. The documents are also relevant to the fair use analysis, particularly for assessing the extent to which OpenAI solicited investments by representing that it would commercialize the at-issue LLMs.

RFP 69 seeks documents “concerning Your policies or practices regarding public access to Your governing documents, financial statements, and conflict of interest rules.” OpenAI objects “to the extent [the request] seeks information and documents that are publicly available.” Are these policies publicly available? If so, please point us to them. The policies and practices (and documents concerning the same) are relevant to fair use, among other issues.

---

<sup>2</sup> Chris Morris, *OpenAI Resignations Are Reaching an Alarming Level. Here Are 11 Key People Who Have Left*, FAST COMPANY (May 17, 2024), <https://www.fastcompany.com/91126785/openai-resignations-are-reaching-an-alarming-level-here-are-11-key-people-who-have-left>.

**b. Requests for Which OpenAI's Responses Are Incomplete or Require Clarification**

RFP 17 seeks documents “concerning the compilation, acquisition, and curation of Training Datasets,” and, by way of example, the RFP identifies three categories of responsive communications. OpenAI asserts that it has “already produced documents concerning the compilation, acquisition, and curation of training datasets.” Has OpenAI already produced documents responsive to all three categories of communications identified in the RFP? Please clarify.

RFPs 20, 21, and 22 seek documents sufficient to show “the total number of records and tokens in Defendants’ Training Datasets” (20), “each source of records in Defendants’ Training Datasets,” (21) and “each article of Times content contained in Defendants’ Training Datasets” (22). For RFP 20, OpenAI agreed to produce documents “sufficient to show the total number of records and tokens in OpenAI’s training datasets”, “if any”, that can be located pursuant to a reasonable search. R&Os at 11. Why did OpenAI qualify its response with the “if any” language? Is OpenAI unsure about whether it has preserved responsive documents? Please also clarify whether OpenAI will produce documents sufficient to show the total number of records and tokens that specifically contain (i) Times Content, (ii) Journalism content, and (iii) copyrighted content. For RFP 21, OpenAI responds that it “already agreed to produce its agreements related to text training data used to train models used for ChatGPT.” R&Os at 12. Are those agreements sufficient to identify all sources of records for the Training Datasets? If not, OpenAI should produce documents sufficient to identify any other sources. Finally for both RFPs 21 and 22, OpenAI points to the forthcoming data inspection process. The Times’s Request for Inspection and Requests for Production serve different purposes. The inspection will permit The Times to inspect data, source code, models, and products; the RFPs seek information and documents in your possession, custody, and control that you bear the burden to search for and produce. Please confirm that OpenAI will produce all responsive documents, regardless of whether the subject matter of those documents relates to The Times’s forthcoming inspection.

RFP 23 seeks documents “concerning all steps taken or contemplated to train, instruct, program, or refine Defendants’ Generative AI Products and Services using Times Content or Journalism content.” OpenAI asserts that it “already produced or made available for inspection . . . documents and text training data for the models used for ChatGPT.” R&Os at 14. That response ignores the portion of the RFP seeking documents about steps “contemplated” to train, as well as the portion addressing Times content or Journalism content. Was that intentional? The response is also vague and incomplete because it is unclear if OpenAI already produced documents concerning the specific aspects of training listed in the RFP, such as finetuning, RLHF, and RLAIIF, as well as any other use of Times or Journalism Content such as for the purposes of grounding these products and services. Please clarify.

RFP 26 seeks documents “concerning Your use or knowledge of web crawlers, bots, spiders, user agents, and related tools to access The Times’s Content.” OpenAI in response points to the forthcoming training data inspection process. Please explain how that inspection process will provide the information requested by this RFP. Specifically, The Times seeks documents related to OpenAI’s policies and practices for third-party robots.txt directives and/or metatags that

block the crawling and scraping of copyrighted content and whether Defendants rely on extending robots.txt permissions surrounding crawling and indexing for Search for GAI purposes. The extent to which GPTBot, Bingbot, and Defendants' other bots respect crawler directives directly impacts the data available to GPT models and the content available for model responses, both of which are clearly implicated in this case.

RFP 27 seeks documents concerning "Your reliance on, use of, or understanding of third-party robots.txt crawler directives or third-party meta tag crawler directives in managing the crawling and indexing activity of your web crawlers, bots, spiders, user agents, GPTBot, any plugins or custom GPTs included on Defendants' platforms, and related applications to access, crawl, and/or index Times Content or third-party content." OpenAI has agreed to produce documents "sufficient to show OpenAI's approach to robot.txt, as explained by [certain] OpenAI blog posts, that it locates pursuant to a reasonable and diligent search." R&Os at 18. OpenAI's "sufficient-to-show" limitation is improper; it should also search for and produce custodial documents because this issue is clearly relevant to how Defendants obtained content to power Defendants' Generative AI Products and Services. The limitation to robot.txt is also improper because the RFP is not so limited.

RFP 28 seeks documents "concerning all measures taken to conceal or obscure Defendants' use of Times Content during any stage of the development of each of Defendants' Generative AI Products and Services." OpenAI has agreed to produce documents "sufficient to show filtering of Times content." R&Os at 19. Please explain what you mean by "filtering." Does "filtering" include measures taken to "conceal or obscure" Defendants' use of Times content? If not, OpenAI should produce those documents as well. OpenAI's sufficient-to-show limitation is also improper, as this is a core issue that directly implicates the willfulness of Defendants' actions (among other issues).

RFP 31 seeks documents "relating to Your technical documentation for Defendants' Generative AI Models, Products, and Services as defined in Annex IV of the European Union's Artificial Intelligence Act." OpenAI responds by identifying "relevant technical documentation" on a particular website. R&Os at 21. Please explain what OpenAI means by "relevant." Does this website have the requested technical documentation, or does it not? Does OpenAI maintain technical documentation sufficient to comply with Annex IV of the European Union's Artificial Intelligence Act? If yes, The Times would expect that documentation to be readily accessible and easily produced.

RFP 32 seeks documents "concerning the training curricula, stages, iterations, cycles, batches, or epochs for Defendants' Generative AI Models." OpenAI improperly limits its commitment to documents "sufficient to show" the requested information. R&Os at 21. OpenAI also claims that it has already produced documents "sufficient to show the training process for the GPT models used for ChatGPT." R&Os at 22. Please identify those documents, and please clarify what you mean by "training process." Does OpenAI intend to exclude other aspects of this RFP, such as "training curricula"?<sup>3</sup>

---

<sup>3</sup> See, e.g., *Language Models Are Few-Shot Learners* at 40, chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://arxiv.org/pdf/2005.14165 (addressing "training curricula").



RFP 36 seeks documents “concerning the differences between Defendants’ Generative AI Models.” OpenAI points to the forthcoming training data inspection process. Is it OpenAI’s position that the Training Datasets are the only aspects of these Models that differ from each other? If not, how will the inspection process provide the requested information?

RFP 43 seeks “surveys, studies, analyses, or statistics on patterns and practices of any of Defendants’ Generative AI Products and Services, including any mentions of The New York Times.” OAI responds that it has “already produced the responsive . . . documents . . . sufficient to show OpenAI’s works that are responsive to this request.” R&Os at 30. The Times does not understand this response. What is an “OpenAI work”? Does that include surveys, studies, analyses, or statistics conducted or generated by OpenAI? Does that include “works” prepared by entities other than OpenAI? Also, please identify the “works” you believe OpenAI has already produced responsive to this request.

RFP 44 seeks documents “concerning any attempts to use Defendants’ Generative AI Products and Services to circumvent paywalls or other access control measures.” OAI responds that it will produce documents “sufficient to show ChatGPT’s approach to paywalls.” R&Os at 31. That sufficient-to-show limitation is improper. Moreover, the request is not limited to seeking information about OpenAI’s “approach” to paywalls. OpenAI should also produce documents, for example, about users’ attempts to evade paywalls through Defendants’ Products and any actions taken by OpenAI in response to such attempts.

RFP 49 seeks documents “concerning Defendants’ monitoring, tracking, and knowledge of investigations into whether Defendants’ AI Models contain copyrighted content, including Times Content.” OpenAI responds that it will produce documents “sufficient to show OpenAI’s knowledge, prior to the inception of this lawsuit, of whether NYT content was included in a dataset for any model used for ChatGPT.” R&Os at 33. Once again, OpenAI’s sufficient-to-show limitation is improper. Furthermore, the RFP is not limited to Times content. It also seeks information about copyrighted content in Defendants’ AI Models more generally. Finally, the timeframe should not be limited to the period before the lawsuit was filed. To the extent OpenAI contends that some post-lawsuit documents would be privileged, OpenAI can withhold and log those documents.

RFP 50 seeks documents “concerning Defendants’ abilities and efforts to track The Times’s use of Defendants’ Generative AI Products and Services.” OpenAI responds that it will produce “responsive, non-privileged policies regarding the use of ChatGPT.” R&Os at 33. The Times does not understand this response. This RFP seeks information about how Defendants track The Times’s use of Defendants’ Products. What are these “policies,” and do they address that issue? If so, how? Do these policies address how Defendants more generally track how users engage with Defendants’ Products, and will you also be producing documents that show how they were or were not applied with respect to The Times? Please clarify.

RFP 60 seeks documents “concerning Defendants’ licensing or use of Bing’s search index.” OpenAI responds that it has not used Bing Index data “to train any GPT model used for ChatGPT and thus will not produce any documents in response to this request.” R&Os at 41. Has OpenAI used Bing Index data to train other Models? Has OpenAI used Bing Index for any other



purpose? Please clarify. OpenAI's position on this RFP is inconsistent with OpenAI's repeated assertions that this case is "very different from the Consolidated Class Action" since this case "focuses heavily on the 'outputs of Defendants' GenAI models.'" Dkt. 72 at 16. OpenAI's position is also inconsistent with OpenAI's agreement "to produce RAG-related documents." Dkt. 147 at 3.

RFP 62 seeks documents "concerning negotiations between Defendants and owners of copyrighted content, including The Times, over Defendants' use of copyrighted content in Defendants' Generative AI Models and Defendants' Generative AI Products or Services." OpenAI responds that "it has already agreed to produce its agreements related to text training data." R&Os at 43. As a threshold matter, those productions are incomplete, as The Times has pointed out in a July 16 email. When will those productions be completed? OpenAI also references the "pending motion to compel regarding at least part of the scope of this request," and says it "will not produce further documents until the court rules on the motion." *Id.* Please clarify that, if The Times prevails on that motion to compel, Dkt. 141, OpenAI will produce the requested documents. The requested documents are also relevant to the fair use analysis, including because these negotiations evidence the existence and nature of a market for the content.

RFP 70 seeks documents "concerning referral, affiliate, advertising, and paid search revenue generated by Defendants' Generative AI Products and Services." OpenAI responds that it will produce documents "sufficient to show income from paid versions of ChatGPT, between its introduction on November 30, 2022 to the present." R&Os at 48. As with many of these responses, OpenAI's "sufficient-to-show" limitation is improper. Moreover, we request a detailed explanation of what is encompassed by "paid versions" of ChatGPT, including any variations between one-time payments, subscriptions, or other monetized features. Additionally, we seek clarification on whether the revenue from advertising, referral, affiliate, and paid search activities, or any other ancillary income related to your Generative AI Products and Services, are included in your response. We also request confirmation that APIs, plugins, and collaborations like Browse with Bing and Copilot are accounted for in the revenue documents. Please also clarify whether OpenAI's response accounts for revenue that OpenAI shares with Microsoft pursuant to any revenue-sharing agreements, or similar agreements with other companies and/or investors. If certain categories of revenue or specific services are being excluded, please provide a justification for these exclusions.

RFP 71 seeks documents "sufficient to show in detail, on a monthly, quarterly, or annual basis, the revenue and profit realized from Defendants' Generative AI Products and Services." OpenAI has agreed to produce documents "sufficient to show income from paid versions of ChatGPT, between its introduction on November 30, 2022 to the present." R&Os at 49. Again, please clarify what you mean by "paid versions" of ChatGPT, including by answering the questions we list for RFP 70, above. In addition, please confirm that you are not seeking to exclude "profit" information from your response, which your response does not mention.

Respectfully,

/s/ Zach Savage  
Zach Savage